Recitation 2: Backpropagation

Recitations

[1] Batched Matrix Multiplication, Broadcasting

[2] Backpropagation

[3] Vectorization + Convolutions

[4] Quantization

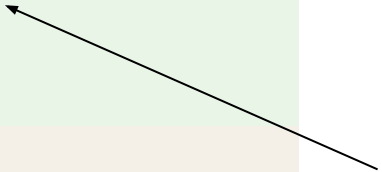
[5] Ethics of AI (Andrew)

[6] Optimizing Attention

[7] TBD

[8] Coding Agents (Ethan)

[9] Deep Learning for Robotics (Bernie)

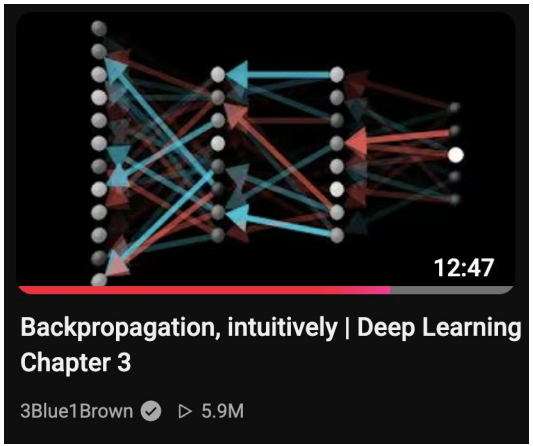


You are here!

Additional practice & Missing prereqs

Hot topics that couldn't fit into the lecture schedule

Personal recommendation for extra reading



[3Blue1Brown on Backpropagation]

[Artem Kirsanov]

Used by all deep learning applications in practise

From-hand / unvectorized, vectorized.

Pytorch, Jax, TensorFlow - autograd frameworks

Lightning, Fabric

Compilation of modules, Activation checkpointing

Chain Rule

Chain Rule for One Independent Variable

Suppose that $x = g(t)$ and $y = h(t)$ are differentiable functions of t and $z = f(x, y)$ is a differentiable function of x and y. Then $z = f(x(t), y(t))$ is a differentiable function of t and

$$\frac{dz}{dt} = \frac{\partial z}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial z}{\partial y} \cdot \frac{dy}{dt}, \tag{14.5.1}$$

where the ordinary derivatives are evaluated at t and the partial derivatives are evaluated at (x, y).

Section 14.5, *Calculus*, Strang and Herman.